NPS-57He7191A

United States
Naval Postgraduate School

THE USE OF A NONADJECTIVAL, NONORDINAL, LINEAR RATING

SCALE IN A SINGLE AXIS COMPENSATORY TRACKING TASK

by

Ronald A. Hess

1 September 1971

This document has been approved for public release
and sale; its distribution is unlimited.

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral A. S. Goodfellow, USN   M. U. Clauser
Superintendent   Academic Dean

ABSTRACT

A brief experimental study is undertaken to determine the utility of a
new pilot rating scale in a fixed base tracking task. The scale is the
nonadjectival, nonordinal, linear scale introduced by C. V. Schufeldt.
The "subcritical" tracking task developed by Jex, McDonnell and Phatak
is utilized in the experiment. The scale's potential for detecting
minor changes in system acceptability is demonstrated.

TABLE OF CONTENTS

I. INTRODUCTION

A.) Background

The Cooper-Harper pilot rating scale for the evaluation of aircraft has found wide acceptance in the field of handling qualities research. The scale, shown in Figure 1, is a means of quantifying a pilot's impressions of the handling qualities of an aircraft which is involved in a specific mission element or task. The scale is adjectival, ordinal and nonlinear in nature. It is adjectival in that descriptors such as "controllable", "adequate", and "satisfactory" appear in the flow diagram used by the pilot. It is ordinal in that handling qualities are ranked in order of decreasing acceptability. It is nonlinear in that a rating of, say 8, does not necessarily indicate handling qualities which are twice as unacceptable as those receiving a rating of 4. The utility of the Cooper-Harper scale has been recently enhanced by a method for predicting ratings[1,2].

As successful and useful as this rating scale has been it is not without its weaknesses. Chief among these are its qualitative character and its ordinal nature. In an attempt to alleviate some of these difficulties, J. D. McDonnell[3] proposed a "global" rating scale for handling qualities investigations. This scale, shown in Figure 2, is an adjectival, nonordinal, linear scale developed through the methods of psychometrics.   While not receiving the wide acceptance of the Cooper-Harper scale, the Global scale has been utilized in handling qualities investigations[4].

1

McDonnell's work centered about finding the coordinates of certain adjectival phrases on a psychological continuum which he called the ψ scale. The adjectival phrases were those most commonly encountered in handling qualities research.

The psychological continuum can be interpreted in the following manner. If a measurement is made on a physical object with a nonhuman instrument of some sort, the measure is an objective one and the resulting data lie along a physical continuum. When a human observer estimates a measure, it is a subjective judgment and the estimates lie along a psychological continuum.

C. V. Schufeldt[5] advanced yet another rating scale. His scale, shown in one of its forms in Figure 3, is nonadjectival, nonordinal, and linear in nature. The impetus behind Schufeldt's research was the idea of developing a scale which would reflect relatively minor differences in system characteristics. To accomplish this, the scale would have to exhibit a good deal of sensitivity without overtaxing the resolution capability of the operator. Schufeldt's hypothesis was that a linear rating scale coincident with the psychological continuum begets such sensitivity. While the Global scale of McDonnell is conceptually close to this realization, Schefeldt felt that in certain applications, the adjectives were a hindrance. He wanted to know if removing the adjectives would allow the rater to transpose his impressions of a system directly to a linear, numerical index. In addition, he wondered if allowing the subject to fractionize his rating would increase scale sensitivity.

Schufeldt investigated his hypothesis by submitting a child's puzzle ("EVEN-STEVEN" by Kohner) to some thirty students in the Department of Aeronautics. Upon successful solution of the puzzle, or at the expiration of an ~~alloted~~ *allotted* time, whichever occurred first, the subject was

asked to rate his impression of the difficulty he encountered in working the puzzle. The subjects indicated their ratings on three different scales, one of which is shown in Figure 3. Schufeldt found a high correlation coefficient (e.g. 0.928 for the scale of Figure 3) between ratings and performance.

B.) Critical-Subcritical Tasks

Encouraged by Schufeldt's results, this author was eager to use the scale in an environment more closely related to handling qualities investigations, i.e. fixed base tracking tasks.

If Schufeldt's scale does indeed possess a sensitivity superior to previous scales, it should yeild better results in areas where these scales were overly sensitive, i.e. the high end (8-10) of the Cooper-Harper scale. If the experiment is to be tractable, the task difficulty should be controlled by as few parameters as possible. Finally, since it was desired to keep the duration of the entire experimental program short, a task which tended to minimize training times should be selected. These criteria pointed toward the selection of the "critical-subcritical" tracking tasks as pioneered by Jex, McDonnell, and Phatak[6].

Critical task (first-order) refers to a special compensatory tracking task in which the real pole, λ, of a first order controlled element

$$Y_c(s) = \frac{\lambda}{s-\lambda}$$

is moved slowly into the right half of the s plane until the subject or operator can no longer maintain control. The value of λ at the onset of instability is called the critical instability score, λ_c. No input is required since operator remnant serves to excite the system[6].

3

Subcritical task (first-order) refers to a similar tracking
situation in which the value of the unstable pole, λ, is kept at a
constant and controllable value, λ_s, throughout the run. In subcritical
tracking, a random appearing input is usually applied. Figure 4 is a
block diagram representing the critical and subcritical systems.

II. EXPERIMENT

A.) Procedure

Fourteen subjects were chosen for the experiment. Of these fourteen,
six were military pilots, two were civilian pilots and six were nonpilots.

The basic experimental procedure went as follows. A subject performed
the critical task experiment twenty times in succession. An average
critical instability score, λ_c, was obtained as the mean of his five
highest λ_c scores. Five subcritical systems were then chosen with pole
locations given by:

$$\lambda_{s_i} = i \cdot \frac{\lambda_c}{6} \qquad i = 1, 2, 3, 4, 5$$

The subject made ten runs of fixed duration, in succession, for each of
these systems. After each set of ten runs, the subject was asked to
rate the system as per the instructions of Figure 5. The five subcritical
systems were ordered randomly and this random order, once selected, was
reversed for every operator. This means operator 1 tracked the subcritical
systems in the order: λ_{s_3}, λ_{s_1}, λ_{s_4}, λ_{s_5}, λ_{s_2}, while for operator 2 the
order was: λ_{s_2}, λ_{s_5}, λ_{s_4}, λ_{s_1}, λ_{s_3}, etc.

The Measurement Systems Inc. isometric, finger grip manipulator
was utilized for the study. The system error was displayed to the
operator as the displacement of a horizontal line on an oscilloscope

screen. The system dynamics, input and mean square error circuits were mechanized on a small analog computer. Table I summarizes the experimental setup. Figure 6 shows the layout.

B.) Discussion

The parameters of Table I were selected to coincide as nearly as possible with those of similar experiments conducted by Systems Technology Inc. (STI)[7]. Due to equipment limitations, the sum of only two sinusoids was used as an input for the subcritical task. Their magnitudes and frequencies were chosen to coincide with those of the two lowest frequency sinusoids used by STI. Were the controlled element, $Y_c(s)$, stable, the sum of just two sinusoids would probably not appear random enough to ensure compensatory behavior. However, the open loop instability made it very difficult for the operator to utilize anything but error information in tracking.

In view of the large number of runs in a single experiment (20 critical + 50 subcritical = 70 runs) it was decided to reduce the subcritical run lengths from an original 100 seconds to 50 seconds. Early experiments with the 100 second lengths resulted in considerable operator fatigue and poor performance. The shorter run lengths, however, probably decreased the accuracy of mean square error scores.

A brief comment on the rating instructions of Figure 5 is in order. At no time was the subject explicitly instructed to associate a particular scale value with a particular system. In addition, each time the subject was asked to evaluate a system, he was given a clean rating sheet.

III. RESULTS

Figure 7 summarizes the experimental results. A set of typical time histories is shown in Figure 8. Table II gives the performance and ratings of the fourteen test subjects. The error scores for the first four subjects were deleted since poor analog scaling caused these values to be inaccurate.

The correlation coefficient for the rating vs. λ/λ_c data is 0.73 as shown in Figure 7. The mean ratings are seen to fall quite close to the regression line. Regression analysis of ratings vs. performance was hampered because of the fact that in five of the subcritical configurations the operators lost control in at least eight of the ten runs. It was difficult to quantify this performance and relate it to that obtained when control was maintained for the full 50 seconds. Hence no further analysis of the error scores beyond that shown in Table II has been presented.

IV. CONCLUSIONS

a.) It does appear that the human operator can transpose his impressions of a system directly to a linear numerical index. The lack of adjectives does not appear to detract from the operator's ability to generate subjective opinion.

b.) The ability of the subject to utilize the linear, nonadjectival scale does not appear to depend upon previous experience with rating

scales in general. The test subjects ranged from the decidedly non-technical (the author's wife) to Navy carrier pilots in the Department of Aeronautics.

c.) The scale appears reasonably sensitive, i.e. the mean ratings are seen to range from 2.9 to 8.4 (55% of the rating scale) as λ/λ_c ranges from 1/6 to 5/6 (66.7% of λ/λ_c scale). The standard deviations of the ratings are fairly uniform across the λ/λ_c scale. This indicates constant sensitivity along the rating scale which is a characteristic of the psychological continuum[3].

It must be emphasized that the rating scale investigated here is not offered as a replacement for the highly successful Cooper-Harper scale. This should be obvious. However, there may arise instances when one desires to detect, in a relative sense, minor changes in system acceptability. In such instances, adjectival scales are simply not appropriate since they lack the necessary sensitivity or overtax the operator's resolution capability. In these cases, a scale such as the one investigated here may prove useful.

REFERENCES

1. Anderson, Ronald O., "A New Approach to the Specification and Evaluation of Flying Qualities," AFFDL-TR-69-120, June 1970, Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio.

2. Dillow, James D., "The 'Paper Pilot' - A Digital Computer Program to Predict Pilot Rating for the Hover Task," AFFDL-TR-70-40, March 1971, Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio.

3. McDonnell, John D., "Pilot Rating Techniques for the Estimation and Evaluation of Handling Qualities," AFFDL-TR-68-76. December 1968, Air Force Flight Dynamics Laboratory, Wright Patterson Air Force Base, Ohio.

4. Stein, Gunter, and Henke, Allen H., "A Design Procedure and Handling Quality Criteria for Lateral-Directional Flight Control Systems," AFFDL-TR-70-152, May 1971, Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio.

5. Schufeldt, C. Vance, "A New Approach to Pilot Rating Scales, M.S. Thesis, August 1971, Department of Aeronautics, U. S. Naval Postgraduate School, Monterey, California.

6. Jex, H. R., McDonnell, J. D., and Phatak, A. V., "A 'Critical' Tracking Task for Man-Machine Research Related to the Operator's Effective Delay Time, Part I: Theory and Experiments with a First-Order Divergent Element," NASA-CR-616, November 1966.

7. Jex, H. R., and Allen, R. W., "Research on a New Human Dynamic Response Test Battery," Proceedings of the Sixth Annual Conference on Manual Control, April 1970, pp. 743-777.
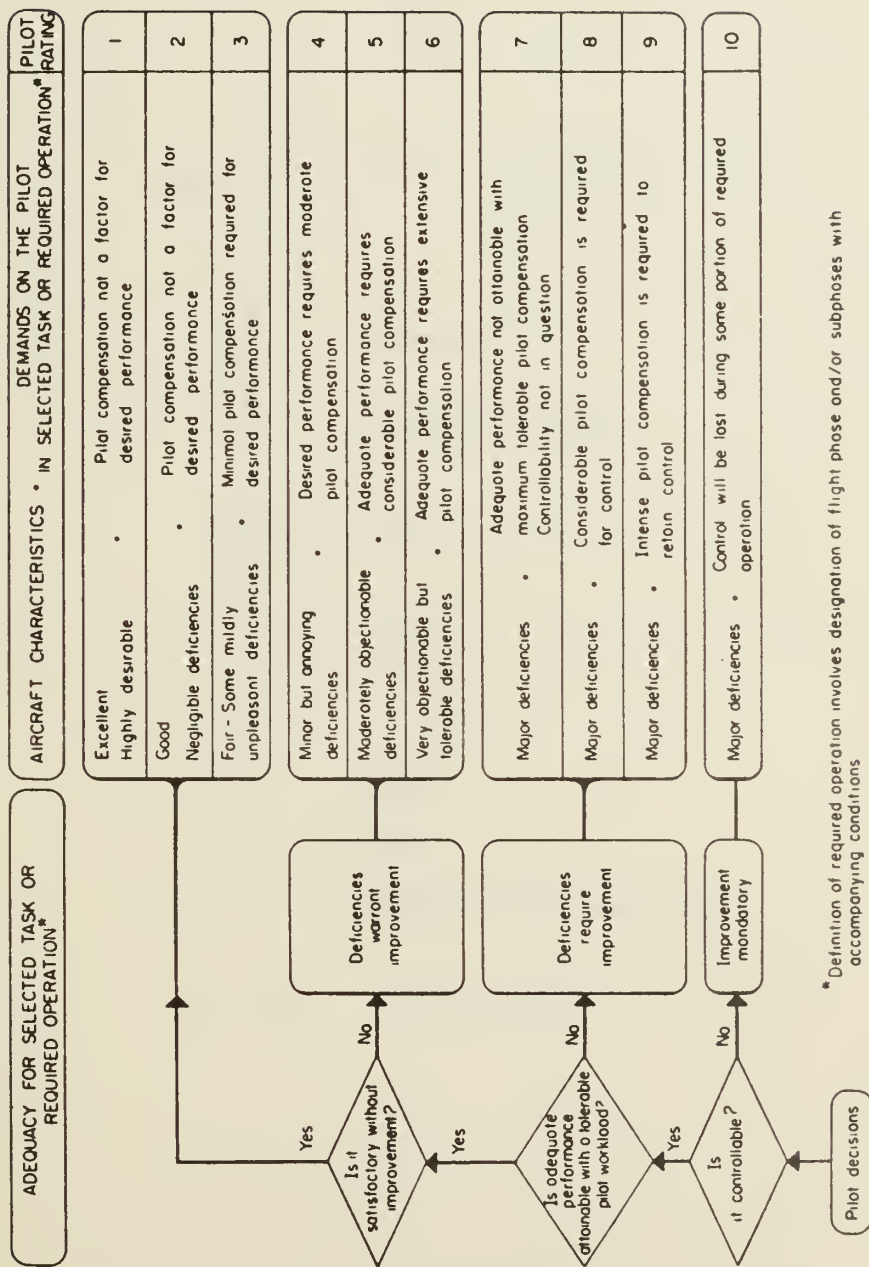
ADEQUACY FOR SELECTED TASK OR REQUIRED OPERATION*	AIRCRAFT CHARACTERISTICS	DEMANDS ON THE PILOT IN SELECTED TASK OR REQUIRED OPERATION*	PILOT RATING
	Excellent Highly desirable	Pilot compensation not a factor for desired performance	1
	Good Negligible deficiencies	Pilot compensation not a factor for desired performance	2
	Fair - Some mildly unpleasant deficiencies	Minimal pilot compensation required for desired performance	3
	Minor but annoying deficiencies	Desired performance requires moderate pilot compensation	4
	Moderately objectionable deficiencies	Adequate performance requires considerable pilot compensation	5
	Very objectionable but tolerable deficiencies	Adequate performance requires extensive pilot compensation	6
	Major deficiencies	Adequate performance not attainable with maximum tolerable pilot compensation Controllability not in question	7
	Major deficiencies	Considerable pilot compensation is required for control	8
	Major deficiencies	Intense pilot compensation is required to retain control	9
	Major deficiencies	Control will be lost during some portion of required operation	10

Deficiencies warrant improvement

Deficiencies require improvement

Improvement mandatory

Is it satisfactory without improvement? Yes / No

Is adequate performance attainable with a tolerable pilot workload? Yes / No

Is it controllable? Yes / No

Pilot decisions

*Definition of required operation involves designation of flight phase and/or subphases with accompanying conditions

Figure 1 Cooper-Harper Rating Scale

9

Favorability of Handling Qualities


```
0 ┌
1 ┝   — Excellent
2 ┝   — Highly Desirable
3 ┝   — Good
4 ┝
    — Fair
5 ┝
    — Poor
6 ┝
    — Bad
7 └   — Nearly Uncontrollable

8 ☐     Uncontrollable
```
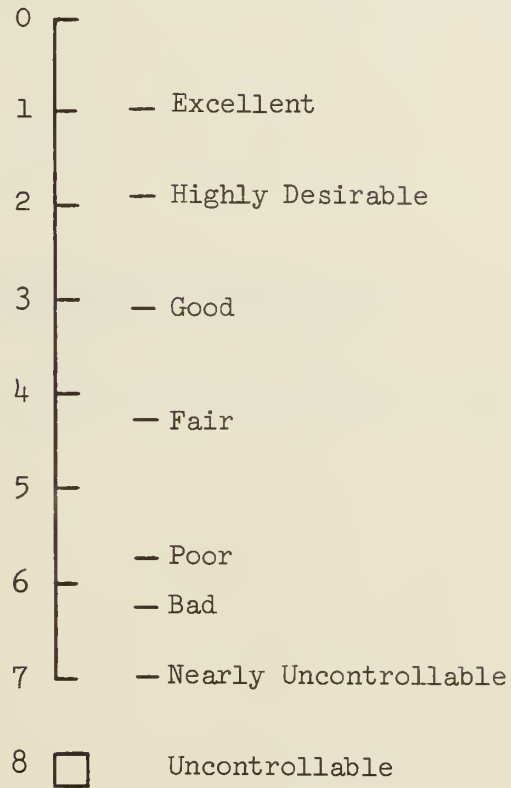
Figure 2 A Global Rating Scale for Handling Qualities Evaluation
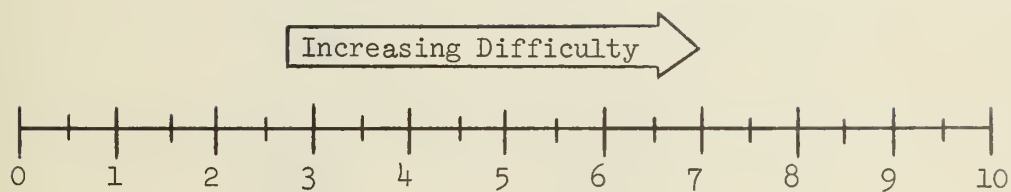
Figure 3 Schufeldt's Nonadjectival, Nonordinal, Linear Rating Scale

Input
i(t)

+

Error
e(t)

−

Display

Displayed
Error

Subject

Control
Force
c(t)

K_c

+

x

$\otimes$

+

λ

λx

$\dfrac{1}{s}$

Output
m(t)

$$Y_c(s) = \frac{K_c \, \lambda}{(s - \lambda)}$$
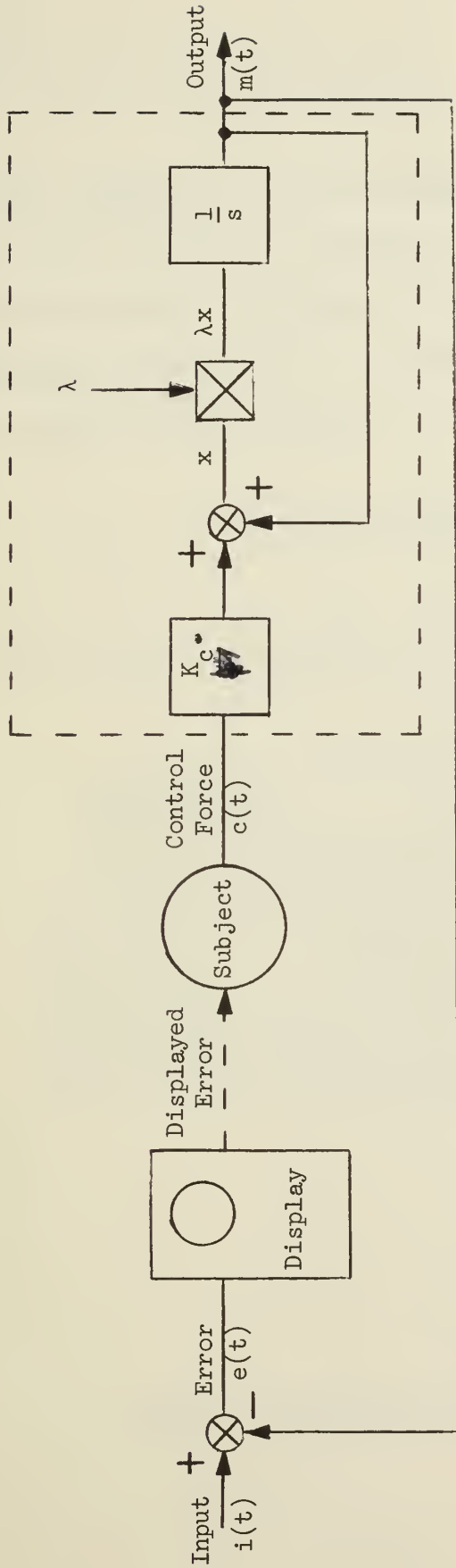
$i(t) = A_1 \sin \omega_1 t + A_2 \sin \omega_2 t$ (Subcritical Task)

$i(t) = 0.0$ (Critical Task)

$\lambda = \lambda_{s_i}$ (Subcritical Task)

$\lambda = \lambda_o + \dot{\lambda} t$ (Critical Task)

Figure 4 Critical and Subcritical Tracking Tasks

12

The **critical** task provided information regarding the limits of your ability to control an unstable system. Using the scale below, indicate the degree of difficulty you encountered in controlling the **subcritical** system checked. All the systems you will be asked to rate in this manner will be unstable.

Increasing Difficulty ⟢

```
0    1    2    3    4    5    6    7    8    9    10
|____|____|____|____|____|____|____|____|____|____|
```

_____ System 1

_____ System 2

_____ System 3

_____ System 4

_____ System 5

Figure 5 Rating Sheet for Subcritical Task
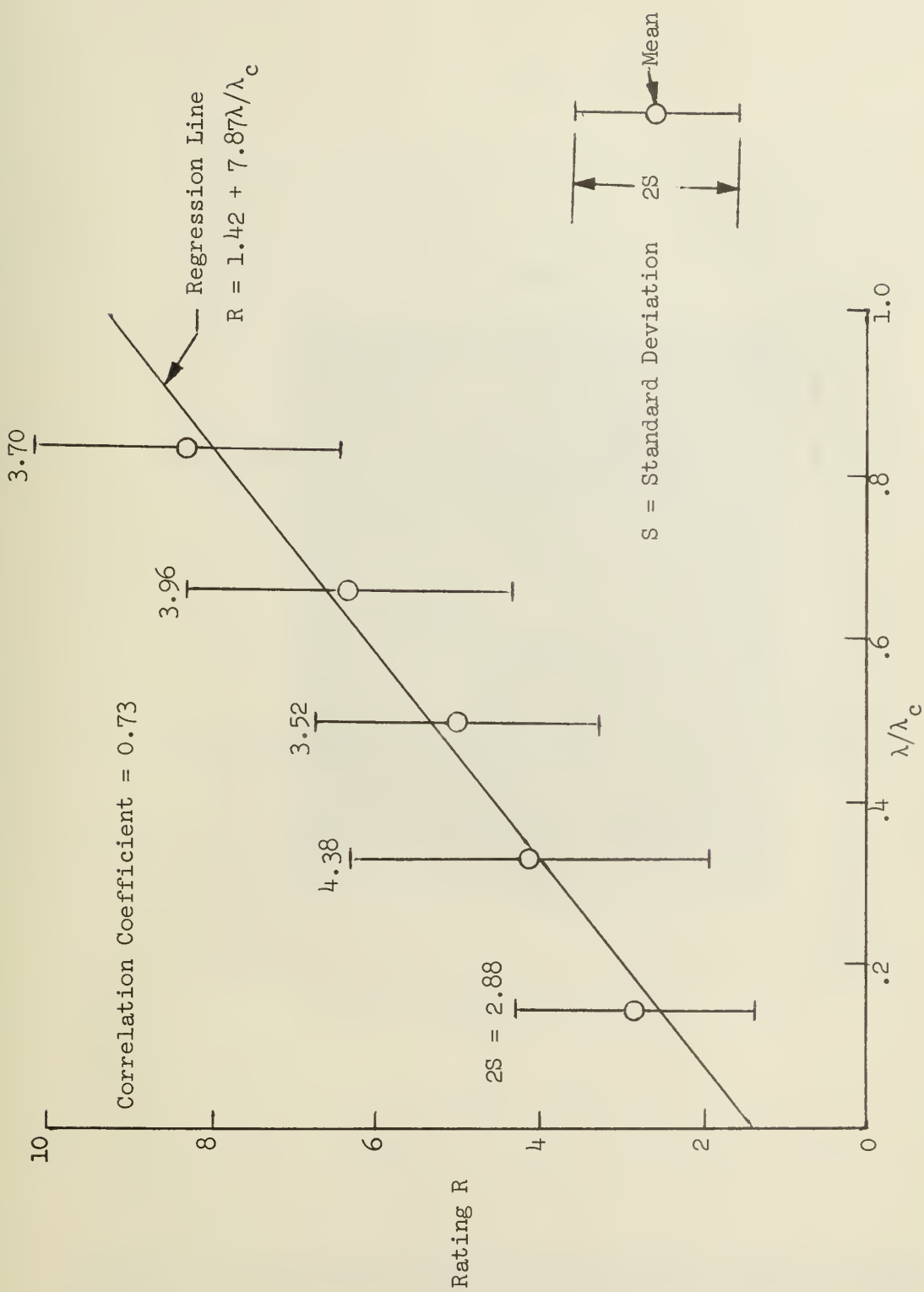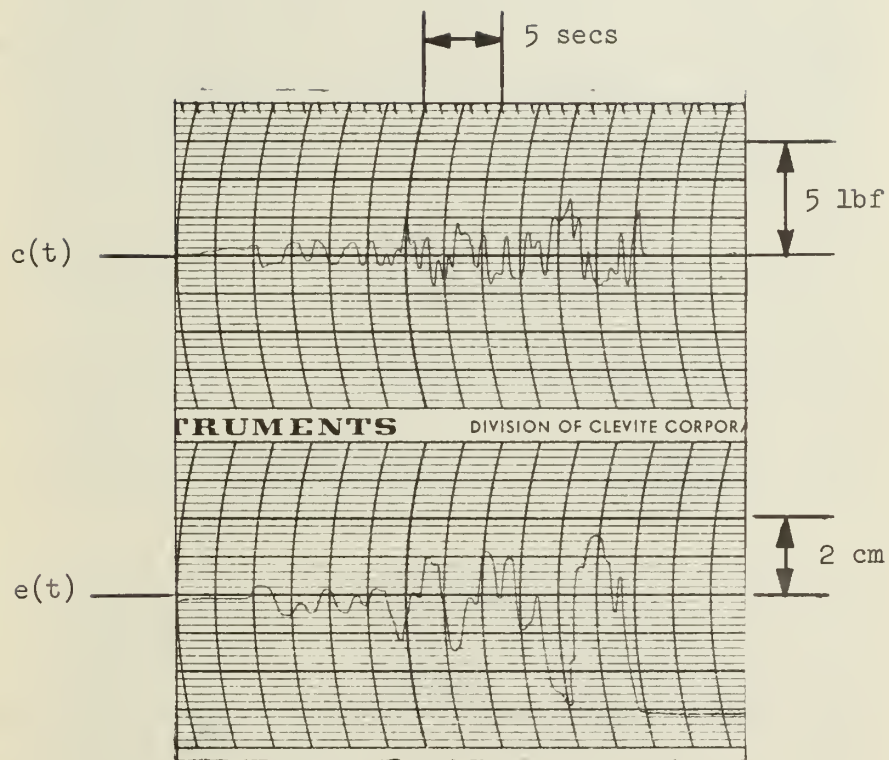
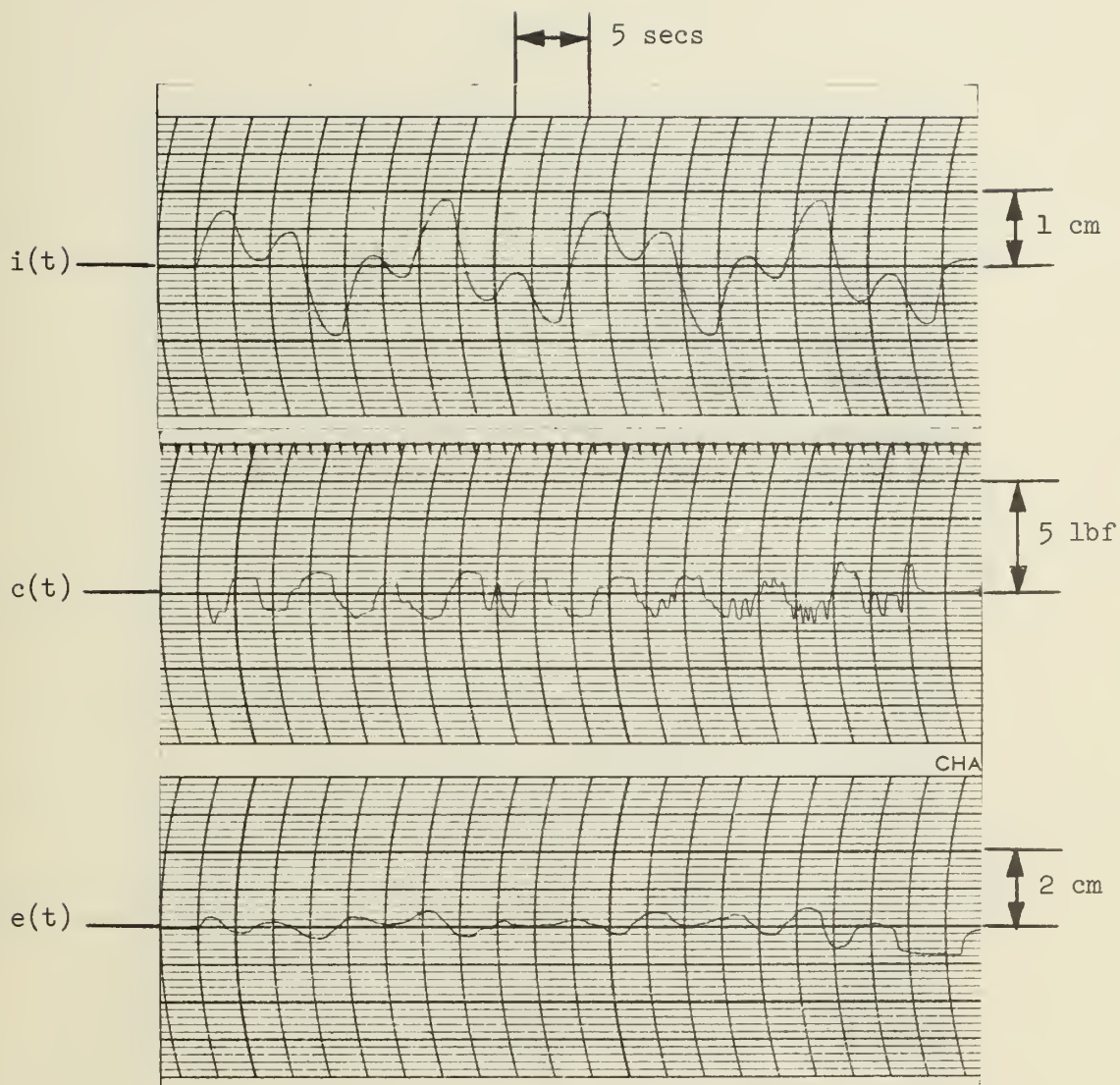Figure 6 Tracking Task Equipment Layout
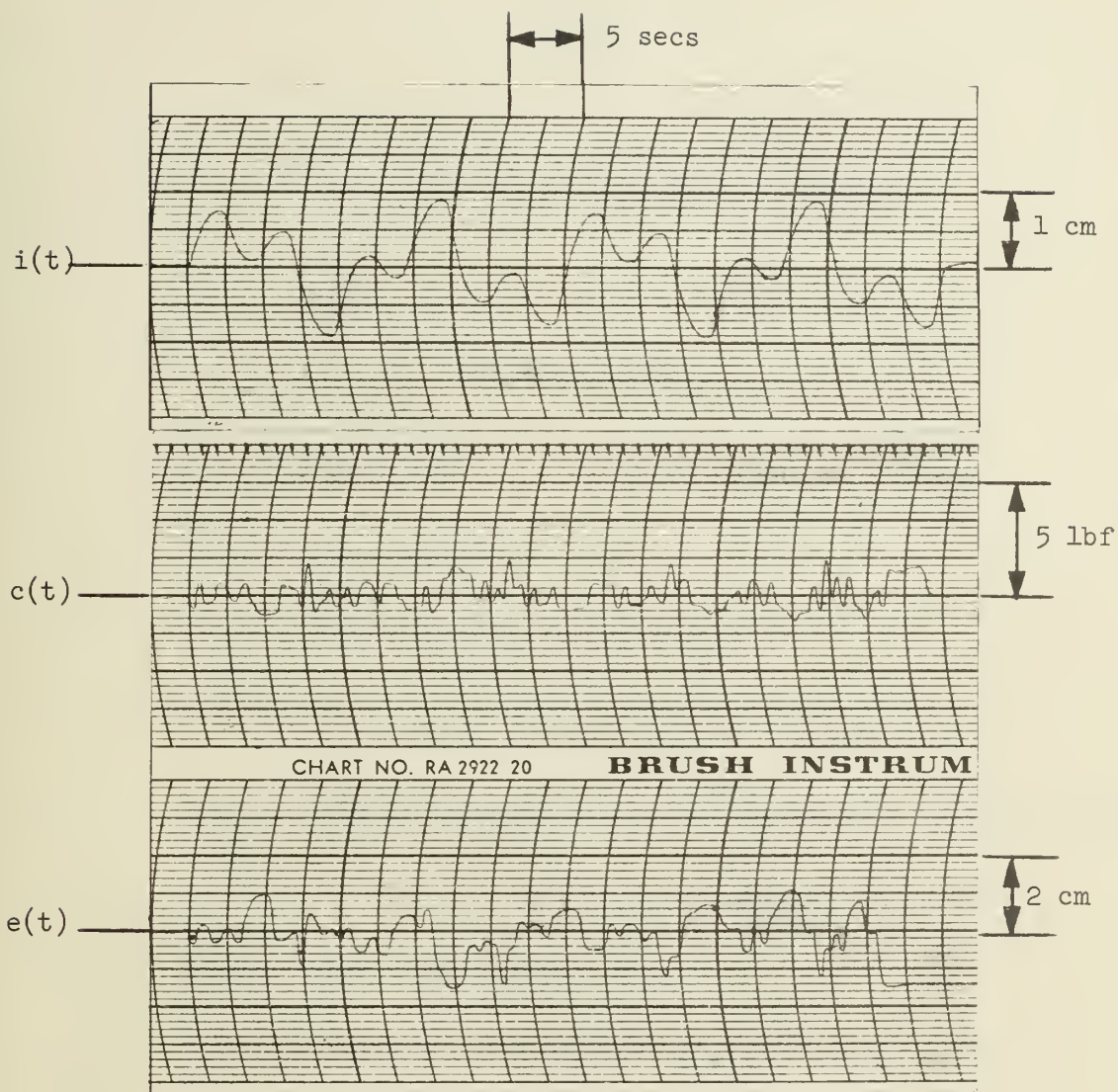
Figure 7 Ratings vs. λ/λ_c

15

Critical Instability Score λ_c = 3.62

Figure 8 Critical Task Stick Output and Error Signals; Subject 14

$\lambda/\lambda_c = 1/6$

Figure 8 cont'd.   Subcritical Task Input, Stick Output and Error Signals;
Subject 14

$$\lambda/\lambda_c = 5/6$$

Figure 8 cont'd.   Subcritical Task Input, Stick Output and Error Signals
Subject 14

TABLE I

Critical and Subcritical Task Parameters

$$Y_c(s) = \frac{\lambda}{s-\lambda}$$

$$\lambda = \lambda_o + \dot{\lambda}t \quad \text{(Critical Task)}$$

$$\lambda_o = 1.0 \text{ rad/sec}$$

$$\dot{\lambda} = 0.1 \text{ rad/sec}^2$$

$$K_c = \text{control/display sensitivity}$$

$$= 0.9 \text{ cm scope deflection/newton stick force}$$

$$K_D = \text{display viewing gain for 50 cm nominal viewing distance}$$

$$= 1.0 \text{ degree visual angle/cm display deflection}$$

$$i(t) = \text{input} \quad \text{(Subcritical Task)}$$

$$= 0.494 \sin 0.502 t + 0.460 \sin 1.256 t \quad \text{cm}$$

$$\overline{i^2(t)} = \text{mean square input}$$

$$= 0.23 \text{ cm}^2$$

TABLE II

Experimental Results - Ratings and Performance

Subj. / Rating $\bar{e}^2/\bar{i}^2$	$\lambda/\lambda_c = 1/6$	2/6	3/6	4/6	5/6	λ_c
1	3.5	5.5	5.0	7.5	9.0	4.77
2	2.0	1.0	5.0	6.5	9.0	4.06
3	0.5	1.0	1.5	3.0	3.5	3.96
4	1.5	2.5	2.1	3.7	6.5	4.18
5	3.0 / .096	4.0 / .167	7.0 / .488	5.5 / .453	10.0 / 1.689	4.47
6	2.5 / .331	4.0 / .270	5.0 / .810	7.0 / 1.215	7.5 / 1.662	3.04
7	2.5 / .410	3.0 / .611	4.5 / .871	8.0 / —	10.0 / —	4.58
8	2.8 / .472	6.5 / .500	4.0 / .993	7.0 / 1.660	9.5 / —	4.28
9	3.3 / .120	1.8 / .182	6.5 / .521	7.0 / .484	8.8 / 1.017	4.23
10	1.5 / .031	4.0 / .117	5.75 / .337	4.5 / .202	7.0 / .821	4.23
11	4.6 / .041	6.1 / .187	5.5 / .406	7.3 / .352	9.2 / 1.802	4.75
12	3.0 / .080	5.0 / .433	4.0 / .193	4.5 / .401	8.0 / —	4.25
13	6.0 / .060	7.0 / .094	7.5 / .392	9.9 / .286	10.0 / .811	4.26
14	4.0 / .286	7.0 / .337	6.0 / 1.050	8.0 / 1.531	9.0 / —	3.81

— Indicates Subject Lost Control in at Least Eight of Ten Runs

INITIAL DISTRIBUTION LIST

No. Copies

1. Defense Documentation Center   20
    Cameron Station
    Alexandria, Virginia 22314

2. Library   2
    Naval Postgraduate School
    Monterey, California 93940

3. Dean of Research Administration   2
    Naval Postgraduate School
    Monterey, California 93940

4. Chairman, Department of Aeronautics   1
    Naval Postgraduate School
    Monterey, California 93940

5. Associate Professor Donald M. Layton   1
    Department of Aeronautics
    Naval Postgraduate School
    Monterey, California 93940

6. LCDR C. Vance Schufeldt, USN   1
    Attack Squadron 212
    Lemoore, California 93245

7. George E. Cooper   1
    Ames Research Center
    Moffett Field, California 94035

8. Robert P. Harper, Jr.   1
    Cornell Aeronautical Laboratory, Inc.
    Buffalo, New York 14221

9. James D. McDonnell   1
    c/o Systems Technology Inc.
    Hawthorne, California 90250

10. Henry R. Jex   1
    Systems Technology Inc.
    Hawthorne, California 90250

11. Bjorn Conrad   1
    Analytical Mechanics Associates
    Palo Alto, California 94302

12. Ronald O. Anderson   2
    Control Criteria Branch AFFDL/FGC
    Air Force Flight Dynamics Laboratory
    Wright Patterson Air Force Base, Ohio 45433

13. Assistant Professor Ronald A. Hess   2
    Department of Aeronautics
    Naval Postgraduate School
    Monterey, California 93940
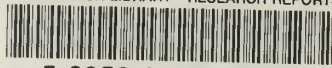
DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Naval Postgraduate School Monterey, California	Unclassified
	2b. GROUP

3. REPORT TITLE

The Use of a Nonadjectival, Nonordinal, Linear Rating Scale in a Single
Axis Compensatory Tracking Task

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report, 1971

5. AUTHOR(S) (First name, middle initial, last name)

Ronald A. Hess

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1 September 1971	27	7

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	NPS-57He7191A
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution
is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

A brief experimental study is undertaken to determine the utility of a new
pilot rating scale in a fixed base tracking task. The scale is the non-
adjectival, nonordinal, linear scale introduced by C. V. Schufeldt. The
"subcritical" tracking task developed by Jex, McDonnell and Phatak is utilized
in the experiment. The scale's potential for detecting minor changes in system
acceptability is demonstrated.

DD FORM 1473 (PAGE 1)
1 NOV 65
S/N 0101-807-6811

23

UNCLASSIFIED
Security Classification
A-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Pilot Rating Scales						

DD FORM 1473 (BACK)
1 NOV 65

S/N 0101-807-6821

24

UNCLASSIFIED

Security Classification

A-31409